Internal Dynamics of a Globular Protein in Water

SHUZO YOSHIOKI

Yatsushiro National College of Technology, Yatsushiro, 866 Japan

Received 27 July 1994; accepted 5 July 1995

ABSTRACT

The frequency distributions of internal dynamics of a protein are calculated in solution using normal mode analysis. Our test case is bovine pancreatic trypsin inhibitor, consisting of 58 amino acid residues. Each water molecule surrounding the protein is treated as an internally rigid body that can move with the vibrating protein. The water molecules are redistributed around the protein, as dictated by the potential energy. It is shown that water molecules around the protein are essential for the protein to keep its tertiary structure close to the X-ray structure. The density of states calculated in this model is shifted toward high frequencies when compared with results previously obtained with a model in which the water molecules were not allowed to move with the protein. This shift toward high-frequency states originates from the stronger interactions of water molecules with the sidechain atoms in the protein. The present model is computationally demanding. So the previous (frozen water) model is suggested to be a reasonable approximation for expressing internal dynamics of a protein in solution.

Introduction

Globular proteins preserve their distinctive tertiary or quaternary structures to make their unique functions possible, but they exhibit a wide range of internal motions, where modes of low frequency less than 200 cm^{-1} are especially important for the biological functions.[1-10] Water molecules around globular proteins are closely related to the stability of the tertiary structures of proteins and to the activation of biological functions. Hence the study of the influence of water molecules on the structure, function, and dynamics of globular proteins is of central importance.[11-17] In this article attention is focused on the influence of water on the tertiary structure and internal

Journal of Computational Chemistry, Vol. 17, No. 7, 878–887 (1996)

CCC 0192-8651 / 96 / 070878-10

dynamics of a globular protein. Bovine pancreatic trypsin inhibitor (BPTI), which consists of 58 amino acid residues, is used as a test case.

In a previous article,[18] a model was presented to explain the internal dynamics of a protein in the frozen solution using normal mode analysis. In the model, hereinafter called model I, water molecules surrounding the protein were assumed to be frozen and the protein was allowed to move in a cavity surrounded by frozen water molecules.

In this article a more realistic model to study the internal dynamics of the protein is proposed. In this model, hereinafter called model II, each water molecule surrounding the protein is treated as a rigid body, but is allowed to move with the protein. This study would be intermediate to a final goal, in which important water–protein coupled vibrations would be found and stretching and bending modes of water molecules surrounding the protein would be found by the normal mode analysis. As independent variables in the present model, internal coordinates in the protein and a translational vector and three rotational angles describing the movement of the protein as a rigid body were chosen, as in model I. In addition, a translational vector and three rotational angles to determine the position of each water molecule were used. With this model II, the normal mode analysis was done for BPTI studied in model I. In model I, the number of the independent geometrical variables was 304 in BPTI plus 6 variables describing translation and rotation. On the other hand, in the present model, there are 304 in BPTI plus 6 variables plus $6 \times N$ for water molecules, where N is the number of water molecules included in the analysis.

Nowadays, with normal mode analysis, the frequency distribution of internal dynamics can be provided rigorously for the protein in the frozen solution assumed in model I. The frequency distribution can also be provided for the protein in a more realistic solution assumed in the present model.

Methods

A fixed laboratory coordinate system (S-system) is used. A protein molecule can move as a rigid body. So we need six variables to specify the position and orientation of the protein with respect to the S-system. In addition, we need six variables for each water molecule to specify the position and orientation of the water molecule with respect to the S-system.

INDEPENDENT VARIABLES DESCRIBING THE PROTEIN

Rotatable Dihedral Angles in the Protein

In the present analysis, bond lengths and bond angles in the protein are treated as fixed, with only rotatable dihedral angles treated as independent variables. Dihedral angles θ_a about bond a in the protein are ordered in a tree structure as follows.[19] The serial number of the rotatable bond increases by one for each step along the tree structure, where no branching occurs. Where branching does occur, the rotatable bond in the shorter branch gets numbered first. The rotatable bond in the other branch gets the number following the number of the last bond in the shorter branch. Each bond connects two units, each of which is a rigid structure consisting of one or more atoms. To each unit we define a local coordinate system. We call the first local coordinate system in the protein the T-system. A conformation of a protein molecule is specified by deciding relative positions and orientations of local coordinate systems in the molecule.

Translational Vector and Eulerian Angles

Here $\boldsymbol{\epsilon}_\nu$ ($\nu = 1,2,3$ or x, y, z) are column unit vectors along positive x, y, z axes of the S-system. The position of the T-system with respect to the S-system is decided by six variables (S-T) specifying relative position and orientation.[19] Thus the origin and the orientation of the T-system with respect to the S-system are designated as $\boldsymbol{t}$ and $\boldsymbol{T}$. Then, if the coordinate vector of an atom with respect to the T-system is $\boldsymbol{r}'$, the coordinate vector of the same atom with respect to the S-system is given by

$$\boldsymbol{r} = \boldsymbol{t} + \boldsymbol{T}\boldsymbol{r}' \tag{1}$$

The S-T variables are specified by three components (t_1, t_2, t_3) in the translational vector $\boldsymbol{t}$ and three Eulerian angles (τ_1, τ_2, τ_3) in the rotation matrix $\boldsymbol{T}$. The rotation matrix $\boldsymbol{T}$, which brings the S-system into the T-system, can be realized as a result of three successive elementary rotations of orthogonal unit axes. Three rotation axes $\boldsymbol{\delta}_\nu$ ($\nu =$

1, 2, 3) are expressed by Eulerian angles (τ_1, τ_2, τ_3), as given by eqs. (7)–(9) in ref. 19.

INDEPENDENT VARIABLES DESCRIBING THE WATER MOLECULES

A coordinate system for each water molecule is defined. The coordinate system in the kth water molecule is the V^k-system. The origin and the orientation of the V^k-system with respect to the S-system are designated as v^k and V^k, respectively. The six variables (S-V) for the kth water molecule are specified by three components (v_1^k, v_2^k, v_3^k) in a translational vector v^k and three Eulerian angles (v_1^k, v_2^k, v_3^k) in a rotation matrix V^k. Three rotation axes $\boldsymbol{\kappa}_\nu^k$ ($\nu = 1, 2, 3$) are also expressed by Eulerian angles (v_1^k, v_2^k, v_3^k), as given in the case of the protein.

ORDER OF INDEPENDENT VARIABLES

The S-T variables precede the tree-structural order of dihedral angles of the protein, and the S-V variables follow the dihedral angles. These variables in this order are designated by q_i [$i = 1, 2, \ldots, n + 6 + (6 \times N)$], where n and N are the number of dihedral angles and water molecules included in the analysis, respectively. In the practical calculation, the coordinate system of the protein is constructed by defining the T-system with respect to the S-system and then by generating local coordinates systems starting from the T-system in the preceding order of the variables. In this method of generation, when a variable θ_a is varied, only a part of the system of the protein moves. This part, which is on one side of the variable θ_a in the tree structure of the system and is on the side not containing the first unit of the protein, is defined as the "moving" side of variable θ_a; it is designated as M_a. The complement of the set M_a is designated as $\overline{M}_a$ (i.e., the "fixed" side of variable θ_a). When a variable is t_ν or τ_ν ($\nu = 1$, 2, or 3), the moving side is designated as M_0 (i.e., the protein). In addition, when a variable is v_ν^k or v_ν^k ($\nu = 1$, 2, or 3), the moving side is designated as M_k (i.e., the kth water molecule).

GRADIENT OF POSITION VECTOR r IN THE PROTEIN OR IN THE WATER MOLECULES

Expressions for the derivatives of the position vector $\boldsymbol{r}_\iota$ for atom ι in the protein, or in the kth water molecule, in the S-system with respect to independent variables (t_ν, τ_ν, θ_a, v_ν^k, v_ν^k) are given by using quantities defined in Table I (first to fourth column) as follows:

$$\frac{\partial \boldsymbol{r}_\iota}{\partial q_i} = \begin{cases} \boldsymbol{\phi}_i \times \boldsymbol{r}_\iota - \boldsymbol{\psi}_i & \text{if } \iota \in M_i \\ 0 & \text{if } \iota \in \overline{M}_i \end{cases} \qquad (2)$$

for $i = 1, 2, \ldots, n + 6 + 6 \times N$. Here $\times$ indicates the cross product.

TOTAL ENERGY E OF PROTEIN–WATER SYSTEM

The protein molecule, with its bond lengths and bond angles fixed, has rotatable dihedral angles as independent variables. As an empirical conformational energy function in protein, the UNICEPP energy function[20] is used as done in model I. So the total energy E for the protein–water system is usually described as a sum of five terms as follows:

$$\begin{aligned} E = & \sum_{\text{dihedrals}} U_{\text{tor}}(\theta_a) + \sum_{ss\ \text{bridge}} U_{\text{loop}}(r_{ij}) \\ & + \sum_{\text{pairs}} \left\{ \epsilon \left(\frac{r_0}{r_{ij}} \right)^{12} - 2.0\epsilon \left(\frac{r_0}{r_{ij}} \right)^{6} \right\} \\ & + \sum_{\text{pairs}} \frac{1}{4\pi\epsilon_0 D} \frac{q_i q_j}{r_{ij}} \\ & + \sum_{h\ \text{bonds}} \left\{ \epsilon \left(\frac{r_0}{r_{\text{HX}}} \right)^{12} - 2.0\epsilon \left(\frac{r_0}{r_{\text{HX}}} \right)^{10} \right\} \end{aligned} \qquad (3)$$

The first term in eq. (3) represents the tortional energy being an explicit function of dihedral angles θ_a in the protein; the second term represents

TABLE I.
Definition of Vectors $\boldsymbol{\phi}_i$ and $\boldsymbol{\psi}_i$, and Diagonal Elements of Hessian.

q_i	$\boldsymbol{\phi}_i$	$\boldsymbol{\psi}_i$	M_i	M_j	s_{ij}	C_{ij}^*
t_ν	0	$-\boldsymbol{\epsilon}_\nu$	M_0	$M_{(k)}$	+1	$C_{\xi\eta}^v$
τ_ν	$\boldsymbol{\delta}_\nu$	$\boldsymbol{\delta}_\nu \times \boldsymbol{t}$	M_0	$M_{(k)}$	+1	$C_{\xi\eta}^v$
θ_a	$\boldsymbol{e}_a$	$\boldsymbol{e}_a \times \boldsymbol{r}_a$	M_a	$\overline{M}_a + M_{(k)}$	+1	$C_{\xi\eta}^v$
v_ν^k	0	$-\boldsymbol{\epsilon}_\nu$	M_k	$M_0 + \overline{M}_k$	+1	$C_{\xi\eta}^v$
v_ν^k	$\boldsymbol{\kappa}_\nu^k$	$\boldsymbol{\kappa}_\nu^k \times \boldsymbol{v}^k$	M_k	$M_0 + \overline{M}_k$	+1	$C_{\xi\eta}^v$

$\boldsymbol{e}_a$ is the unit vector along the bond a pointing toward the direction of its moving side, and $\boldsymbol{r}_a$ is the position vector of an atom at either end of bond a. The set $M_{(k)}$ includes all atoms in the water molecules included in the analysis, and $\overline{M}_k$ is the complement of the set M_k (i.e., it includes all atoms in the water molecules except for the kth water molecule). $C_{\xi\eta}^v = C_{\eta\xi}^v$.

the cystine loop-closing energy being a function of a distance r_{ij} between a pair of atoms i and j in disulfide bonds. The third term represents the nonbonded interaction energy. These interactions take place between all nonbonded atoms in the system. The parameters ϵ and r_0 are given in the UNICEPP program. The fourth term represents the electrostatic interaction energy. D is the relative dielectric constant. The fifth term represents the energy of hydrogen bonds in the system. These interactions are summed only over atoms which can participate in hydrogen bonds. The term r_{HX} is a distance between a donor (H) and an acceptor (X). As in the nonbonded energy, values of ϵ and r_0 for each donor–acceptor combination are given in the UNICEPP program.

Our concern is to calculate the first and second derivatives of E with respect to variable q_i. The derivatives of the first term of eq. (3) with respect to θ_a are trivial. So, in the following, the trivial terms of eq. (3) are not considered explicitly. The first and second derivatives of the first term of eq. (3) are to be added to the results obtained in the following for the remaining terms of eq. (3).

FIRST DERIVATIVES OF TOTAL ENERGY *E*

With eq. (2), the first derivatives of the total energy E of the protein–water system with respect to q_i can be summarized into one equation by using quantities defined in Table I (first to fifth column) as follows:

$$\frac{\partial E}{\partial q_i} = (\boldsymbol{\phi}_i, \boldsymbol{\psi}_i) \cdot (-1) \sum_{\xi \in M_i} \sum_{\eta \in M_j} c_{\xi\eta} \begin{pmatrix} \boldsymbol{r}_\xi \times \boldsymbol{r}_\eta \\ \boldsymbol{r}_\xi - \boldsymbol{r}_\eta \end{pmatrix} \quad (4)$$

Here $c_{\xi\eta}$ is given by eq. (17) in ref. 19. The first and the last factors on the right-hand side in eq. (4) are six-dimensional row and column vectors, respectively.

HESSIAN MATRIX FOR TOTAL ENERGY *E*

The second derivative $\partial^2 E / \partial q_i \, \partial q_j$ for the total energy E of the protein–water system can be summarized into a single formula as follows:

$$\frac{\partial^2 E}{\partial q_i \, \partial q_j} = (\boldsymbol{\phi}_i, \boldsymbol{\psi}_i) \times \left\{ s_{ij} \sum_{\xi \in M_i} \sum_{\eta \in M_j} (c_{\xi\eta} \boldsymbol{C}_{ij}^* + d_{\xi\eta} \boldsymbol{D}_{\xi\eta}) \right\} \begin{pmatrix} \boldsymbol{\phi}_j \\ \boldsymbol{\psi}_j \end{pmatrix} \quad (5)$$

Here $d_{\xi\eta}$, $\boldsymbol{C}_{\xi\eta}$, and $\boldsymbol{D}_{\xi\eta}$ are given by eqs. (22), (26), and (27) in ref. 19, respectively. Only the diagonal elements of the Hessian are given in Table I.

CONFIGURATION OF WATER MOLECULES

The configuration of water molecules surrounding the protein was taken to be the first water configuration in environment b, as in earlier work.[18] This water configuration is briefly reviewed.

First, the protein was put into a rectangular box and water molecules were added to fill the remaining empty space. The box size was selected so that the shortest distance between any atom of the protein and one of the walls of the box was 8 Å. After filling the empty space, water molecules within 2.3 Å from any atom on the surface of the protein were eliminated and 3096 water molecules remained. Of the 3096 water molecules, 219 were retained in a shell between 2.3 Å and 4.0 Å from the protein surface; each of these solvent molecules had a negative interaction energy with the protein. For the protein to form favorable interactions with neighboring water molecules, the whole system of protein and internally fixed water molecules was energetically relaxed by the method of steepest descents.[21] The conformation of this solvated protein–water was used as the starting one, which is shown in Figure 2b of ref. 18.

ENERGY FUNCTION

In the present work (model II), the same energy function as in model I[18] for the intraprotein and protein–water interactions was used; the value of energy parameter ϵ in the third and fifth terms in eq. (3) was reduced to one fifth for the protein–water interaction, and Scheraga atom types[20] 18 and 4 for O and H of the water molecule, respectively, were chosen in the calculation of nonbonded and hydrogen bond energy in eq. (3). In the calculation of the electrostatic energy for the protein–water interaction, the partial charges of O and H in water were chosen as -0.330 and 0.165, respectively, in electronic charge units. The dielectric constant D is 2.

In calculation of the water–water interaction energy, the value of ϵ in the third and fifth terms in eq. (3) was not reduced, and the dielectric constant D is 1. The other values in the third, fourth, and fifth terms were the same as in model I.

No truncation schemes were used in the energy calculation for the intraprotein interactions, protein–water interactions, and water–water interactions.

ENERGY MINIMIZATION

The system was further energy minimized using Newton's method based on a modified Cholesky factorization of the Hessian.[22] Energy minimization was stopped at a point in multidimensional space $\{t_\nu, \tau_\nu, \theta_a, v_\nu^k, v_\nu^k\}$, at which the root mean square (rms) energy gradient of $\partial E/\partial\theta_a$ only for the protein was less than 0.1 (kcal/mol/degree) and the subset of the Hessian, whose elements are consisted of $\partial^2 E/\partial\theta_a\partial\theta_b$, had all positive eigenvalues. At this point, the rms energy gradients of $\partial E/\partial v_\nu$ and of $\partial E/\partial v_\nu$ for water molecules were less than 0.01 (kcal/mol/Å) and 0.001 (kcal/mol/degree), respectively. Hence it seems acceptable to use only the gradients of the protein as a stopping criterion in the minimization. Because a sufficiently small rms gradient could be obtained, the protein–water conformation appears to be close to a minimum. The final protein–water system is called conformation M_{II}.

Results

ENERGY MINIMUM CONFORMATION

The water molecules distributed uniformly around the protein in the starting conformation moved in energetically favorable sites in the minimum conformation M_{II}. The rms deviation of the C^α atom for conformation M_{II} from the X-ray structure was 0.255 Å, and that for all atoms in the protein was 0.515 Å. Figure 1 shows that conformation M_{II} fits nicely on the X-ray structure, although the places of the water molecules, not shown here, are different from the starting places.

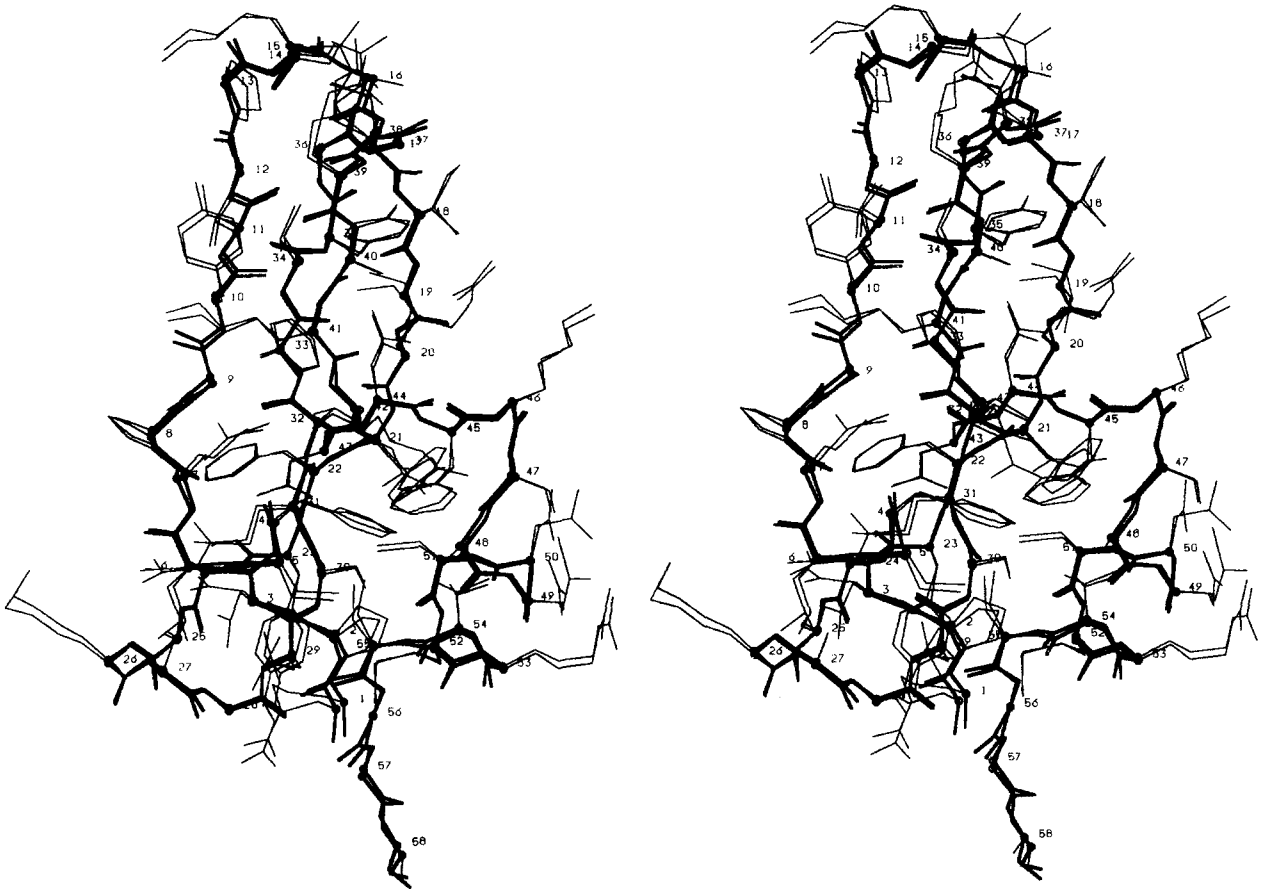

FIGURE 1. Stereoview of the protein in conformation M_{II} best superposed on the X-ray structure. Residues are numbered at positions of C^α atoms in the X-ray structure.

Is the protein conformation obtained in this model trapped on a local minimum? Using as a criterion of the global minimum a small value of rms deviation from the observed X-ray structure, the protein conformation obtained here would be almost so, as shown in the rms deviation calculated earlier and in Figure 1. The small rms deviation is apparently derived from the fact that the environment of the protein interacting with surrounding water molecules resembles the solvated nature of the protein in its crystalline environment. On the other hand, the energy minimum conformation of the protein in vacuo has a large rms deviation from the X-ray structure (i.e., 1.322 Å for C^{α} atoms), and we can see a stereoview of the minimum conformation best superposed on the X-ray structure, which displays large deviations between the two structures (see Fig. 4c in ref. 18). It is well known that energy minimization of a "dry" protein will make it overly compact with respect to what is seen in the crystalline state. (Crystallized proteins typically have a high percentage of water.) Water molecules around the protein are essential for the protein to maintain the tertiary structure necessary for biological activity. The question of whether an in vacuo minimum structure deviating from an X-ray structure should be treated as native or denatured needs to be investigated experimentally.

DENSITY OF STATES

Using the Hessian for the protein with all positive eigenvalues, normal mode analysis was done. The calculated frequencies of vibrational normal modes were convoluted with the instrumental resolution function (see Fig. 3 in ref. 11), and the density of states is given in Figure 2. To compare the density of states obtained in this model II with the previous results on model I and for an in vacuo environment, the densities of states for conformation b and in vacuo conformation (see ref. 18) are shown (Fig. 2). We see that the density of states in this model II is similar in form to the one in model I, but the density of states is shifted toward high frequencies in comparison to the result in model I. This is caused by water molecules bound mainly to sidechain atoms in the protein. In

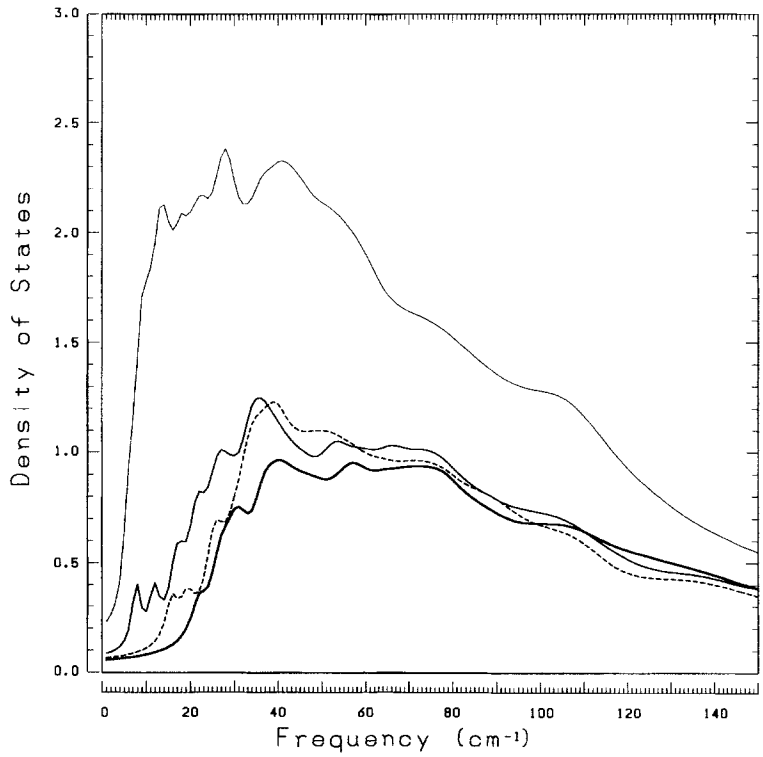

FIGURE 2. The density of states calculated from normal mode analysis and convoluted with an instrumental resolution function. The units are modes per wave number. The dashed curve corresponds to conformation $M_I(\epsilon/2)$, discussed later. The top curve corresponds to the in vacuo conformation. The middle curve corresponds to conformation b. The bottom curve corresponds to conformation M_{II}.

model I, even if water molecules experienced attractive potential energies, they were not allowed to move to preferred sites. So the void surrounded by water molecules in model I might be a little large. The density of states for the in vacuo conformation trapped on a local minimum shows the larger density of states in the low-frequency range, and then it brings out the larger fluctuations than in model II.

ATOMIC POSITION FLUCTUATIONS

In Figure 3, atomic position fluctuations of the C^{α} and C^{β} atoms, averaged over those of each residue that occur when normal modes are excited thermally at 300 K, are plotted against residue number for conformation M_{II} in model II and conformation b in model I. This figure also shows atomic position fluctuations of sidechain atoms.

Discussion

CPU TIME OF ENERGY MINIMIZATION

In model II, the number of the independent variables for the protein–water system is $1624 = 6 + 304 + (6 \times 219)$. Energy minimization to get conformation M_{II} required 67.5 h of CPU time on the FACOM-M1800 computer in Kyushu University. On the other hand, in model I, the number of the independent variables for the protein is $310 = 6 + 304$. Energy minimization to obtain conformation b required 1.7 h of CPU time on the same computer. The ratio of the number of variables in model II to model I is 5.2, and the ratio of CPU time is 40, which is more than the square of the ratio of the number of variable. The preceding ratio shows that the computations in model I are far less demanding.

In the present work, Newton's method based on a modified Cholesky factorization of the Hessian was used. Some approaches to minimization that were tried were as follows. The Hessian was resolved into three subsets; the first subset consists of only a translational vector and Eulerian angles of the protein, the second subset consists of only the dihedral angles in the protein, and the third subset consists of only the variables of water molecules. With the modified Cholesky factorizations of each subset, the line search toward a minimum conformation was performed. The second approach was that the Hessian was resolved into two subsets; the first subset consists of only a translational vector, Eulerian angles, and the dihedral angles of the protein, and the second subset consists of only the variables of water molecules. The line search was also done as in the first approach. The third approach was that the whole Hessian was used for the line search. With a combination of these three approaches, the minimiza-

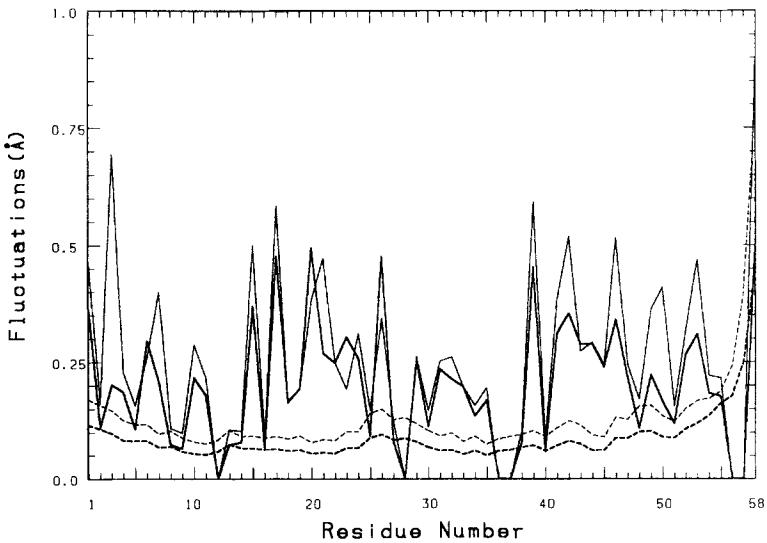

FIGURE 3. Amplitudes of atomic position fluctuations are plotted against residue number. The lower two curves correspond to amplitudes of C^{α} and C^{β} atoms, averaged over those of each residue, whereas the upper two curves correspond to amplitudes of sidechain atoms, averaged over those of each residue. The thin curves in the both of lower and upper curves correspond to conformation b, and thick curves correspond to conformation M_{II}.

tion was done as described here. In the beginning, the first approach was adopted to reduce the energy of the protein–water system for the duration of about 3 h of CPU time; the reduction in energy in this CPU time was 52% of the total reduction of energy. Hence the energy reduction rate for CPU time in the first approach was 286.23 (kcal/mol/h). After this, the third, second, and first approaches were adopted following this order. As a result, three energy reduction rates were 13.22, 14.33, and 0.65 (kcal/mol/h) for the first, second, and third approaches, respectively.

At the beginning of the minimization, the first approach was efficient. Although the second was not used at the beginning of the minimization in this calculation, the second approach would probably also be efficient because the first and second approaches had almost the same energy reduction rates in the intermediate and last stages of the minimization. On the other hand, the third approach was not efficient. This indicates that the step sizes of all the variables toward a minimum conformation were small compared to the other two approaches, because the modified Cholesky factorization was done for the Hessian expressed by all the variables. The third approach, however, might sometimes be effective for a line search in a complicated energy space.

COMPARISON OF MODELS I AND II

As shown in Figure 2, the density of states in model II is shifted to high frequencies compared to the result in model I. Reflecting this shift, magnitudes of atomic position fluctuations of the C^{α} and C^{β} atoms, almost buried in the protein, for model II are reduced compared to the results for model I (Fig. 3). As for fluctuations of the sidechain atoms, the reductions in the magnitude of fluctuations for model II in comparison to model I are different for each residue, reflecting the binding of water molecules to sidechain atoms.

To express how a normal mode exhibits movements of the protein, displacement vectors of the normal mode were examined. The normal mode with frequency of 21.90 cm^{-1} shows that displacement vectors of Gly 57 and Ala 58 are predominant. In addition, the normal mode with frequency of 26.51 cm^{-1} shows that those of Arg 20 are also significant. Although the latter frequency is the second lowest, motions of atoms are localized in one sidechain. This behavior is different from the one in model I for the normal modes with the low frequencies. The minimum conformation M_{II} shows that some water molecules departed from near the C terminus and around the sidechain of Arg 20. Depletion of water molecules will affect the localized motions of the protein atoms. Therefore, it is necessary to surround the protein with more water molecules to analyze the protein–water system more qualitatively and quantitatively.

The number of water molecules necessary for adequate analysis cannot be estimated. The analysis, including more water molecules in model II, demands CPU time greater than 67.5 h, although the CPU time could be shortened by applications of both the first and second approaches in the minimization process. If we were to include 100 water molecules, for example, the ratio of the number of variables in model II to model I would be 7.1 and the CPU time needed would be greater than 86 h (1.7×7.1^2). The calculation itself is expensive. As an alternative, application of model I with a stronger interaction energy between protein and water molecules to imitate internal dynamics of the protein in model II can be suggested.

APPLICATION OF MODEL I

Mean square amplitudes of atomic fluctuations for C^{α} and C^{β} atoms in models I and II are 0.027 $Å^2$ and 0.011 $Å^2$, respectively. The ratio of this value in model I to model II is 2.5. The smaller mean square amplitude in model II compared to model I originates from the stronger interactions of water molecules with sidechains in the protein. To mimic the protein in model II, the energy parameter $\epsilon/5$ between protein and water molecules used in model II was increased to $\epsilon/2$. The new minimized conformation determined for model I is called conformation $M_I(\epsilon/2)$. Because the water molecules are not allowed to move with the protein, they are uniformly distributed around the surface of conformation $M_I(\epsilon/2)$. The calculated density of states is given in Figure 2. We see that the density of states in conformation $M_I(\epsilon/2)$ is shifted toward high frequencies in comparison to the result in conformation b of model I, reflecting the stronger interactions between the protein and water molecules. However, the shape of the density of states in conformation $M_I(\epsilon/2)$ is different from the one in conformation M_{II} in the frequency range 30 cm^{-1} to 60 cm^{-1}. In addition, the magnitudes of atomic position fluctuations in conformation $M_I(\epsilon/2)$ are intermediate between the results in conformation b and in conformation M_{II} (not shown here). The larger fluctuations in conformation $M_I(\epsilon/2)$ compared to conformation M_{II} derive

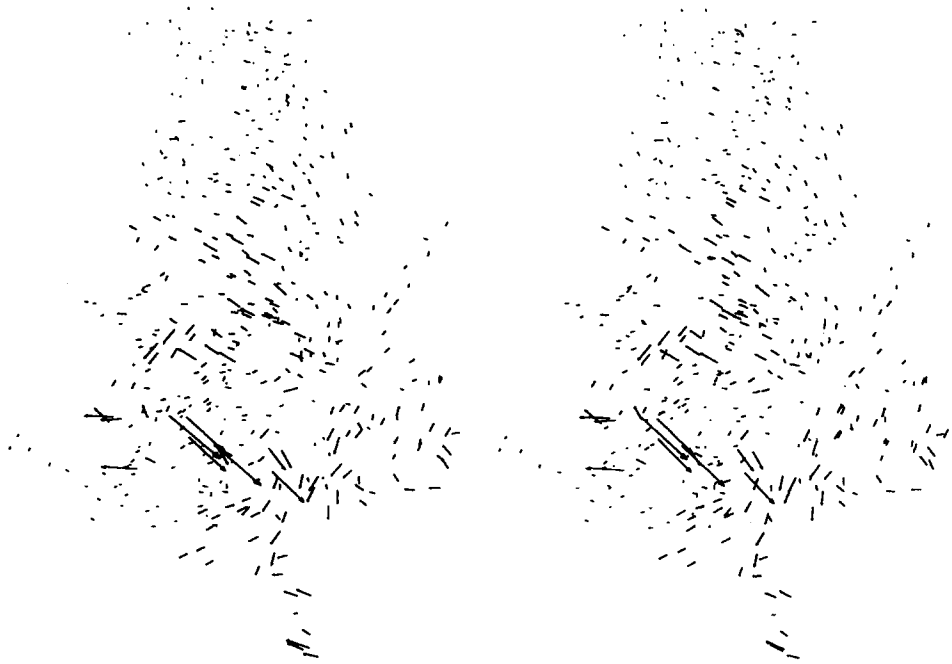

FIGURE 4. Stereoview of displacement vectors of atoms in conformation $M_I(\epsilon)$ in normal mode of the second lowest frequency 23.70 cm^{-1}. Displacements of atoms that occur when this mode is excited thermally at 300 K are magnified 20 times for easy perception.

from the density of states in the frequency range 30 cm^{-1} to 60 cm^{-1}.

It would be possible to reduce the magnitudes of atomic fluctuations in conformation $M_I(\epsilon/2)$ to results in conformation M_{II} by further intensifying the interaction energy of the protein with the water molecules. For instance, increasing the energy parameter $\epsilon/2$ to ϵ might yield magnitudes of atomic position fluctuations comparable to those in conformation M_{II}. The analyses using the energy parameter ϵ had been already done in the previous article (see Fig. 10 in ref. 18). Call the conformation already obtained conformation $M_I(\epsilon)$. Atomic position fluctuations in $M_I(\epsilon)$ (not shown) are comparable to those in conformation M_{II}, although the atomic fluctuations in the sidechains are a little different from the results in conformation M_{II}. In addition, the displacement vectors of the normal modes with the low frequencies are not localized in one sidechain (see Fig. 4) because of the uniform distribution of water molecules around conformation $M_I(\epsilon)$.

Thus the application of model I to express the internal dynamics of protein in solution could be considered a good approximation to model II. In the case of BPTI, the frozen water molecules surrounding the protein were able to mimic dynamics of the protein in model II by model I. Hence the method adopted in model I to build the frozen water molecules around the protein appears an adequate approximation if the energy parameter ϵ is suitably adjusted.

References

1. I. D. Kuntz and W. Kauzmann, *Adv. Prot. Chem.*, **28,** 239 (1973).
2. J. T. Edsall and H. A. Mackenzie, *Adv. Biophys.*, **10,** 137 (1978).
3. J. T. Edsall and H. A. Mackenzie, *Adv. Biophys.*, **16,** 53 (1983).
4. J. A. Rupley, P. H. Yang, and G. Tollin, In *Water in Polymers,* ACS Symposium 127, S. P. Rowland, Ed., American Chemical Society, 1980, p. 111.
5. J. L. Finney, J. M. Goodfellow, and P. L. Poole, In *Structural Molecular Biology,* D. B. Davies, S. Danyluk, and W. Saenger, Eds., Plenum, New York, 1982, p. 387.
6. B. Jacrot, S. Cusack, A. J. Dianoux, and D. M. Engelman, *Nature,* **300,** 84 (1982).
7. P. L. Poole and J. L. Finney, *Biopolymers,* **22,** 255 (1983).
8. P. L. Poole and J. L. Finney, *Biopolymers,* **23,** 1647 (1984).
9. J. Smith, S. Cusack, P. Poole, and J. Finney, *J. Biomol. Struct. Dynam.*, **4**(4), 583 (1987).
10. B. F. Rasmussen, A. M. Stock, D. Ringe, and G. A. Petsko, *Nature,* **357,** 423 (1992).
11. J. Smith, S. Cusack, U. Pezzeca, B. Brooks, and M. Karplus, *J. Chem. Phys.*, **85,** 3636 (1986).

12. M. Levitt and R. Sharon, *Proc. Natl. Acad. Sci. USA*, **85,** 7557 (1988).
13. J. Smith, K. Kuczera, B. Tidor, W. Doster, S. Cusack, and M. Karplus, *Physica B*, **156 & 157,** 437 (1989).
14. R. J. Loncharich and B. R. Brooks, *J. Mol. Biol.*, **215,** 439 (1990).
15. J. Smith, S. Cusack, B. Tidor, and M. Karplus, *J. Chem. Phys.*, **93,** 2974 (1990).
16. A. Kitao, F. Hirata, and N. Gō, *Chem. Phys.*, **158,** 447 (1991).
17. S. Hayward, A. Kitao, F. Hirata, and N. Gō, *J. Mol. Biol.*, **234,** 1207 (1993).
18. S. Yoshioki, *J. Comput. Chem.*, **15,** 684 (1994).
19. W. Braun, S. Yoshioki, and N. Gō, *J. Phys. Soc. Jpn.*, **53,** 3269 (1984).
20. F. A. Momany, L. M. Carruthers, R. F. McGuire, and H. A. Scheraga, *J. Phys. Chem.*, **78,** 1595 (1974).
21. W. F. van Gunsteren and H. J. C. Berendsen, Groningen Molecular Simulation (GROMOS) Library, BIOMOS, Nijenborgh 16, 9747 AG Groningen, The Netherlands, 1987.
22. P. E. Gill, W. Murray, and M. H. Wright, *Practical Optimization*, Academic Press, London, 1981, p. 105.